The 12th International Conference on
Computer Science & Education (ICCSE 2017)
August 22-25, 2017. University of Houston, USA

Effective Dynamic Algorithms for Massive Mark Point Aggregation Display

Kunhui Lin
Software School of Xiamen University
Xiamen, China
khlin@xmu.edu.cn

Yangbin Pan
Software School of Xiamen University
Xiamen, China
1004567472@qq.com

Xiaoli Wang*
Software School of Xiamen University
Xiamen, China
xlwang@xmu.edu.cn

Mengsang Wu
Software School of Xiamen University
Xiamen, China
423848331@qq.com

Shiyu Su
Software School of Xiamen University
Xiamen, China
sysu1993@gmail.com

***Abstract*—In the GIS applications, massive mark point display has become a very important problem. Especially for the application in vehicle networking, high-load vehicle position moving increases the difficulty of the display problem. Many existing efforts have been taken on proposing aggregation algorithms for map mark points. However, all these works suffer from efficiency and scalability problems when we employ them to the application in the vehicle networking. In order to solve the problem of high efficiency map display of high load vehicle location data, this paper proposes the dynamic aggregation algorithm based on the administrative unit and the dynamic aggregation algorithm for K-means service mark point based on the administrative unit. Experimental results show that these two algorithms can make the location data of the vehicle highly efficient display on the map, and can achieve smooth interaction and friendly user experience.**

***Index Terms*—vehicle networking, mark point, dynamic aggregation algorithm, k-means.**

I. INTRODUCTION

With the development of the internet of things and the increasing number of vehicles, the research of vehicle networking technology has become a trend [1][2][3]. How to accurately display the positions of vehicles, especially highway vehicles on the map is the core problem of vehicle networking. Displaying a large amount of data on the map, not only makes the map appear very messy, but also takes up a lot of system resources, which cause the browser to crash and bad user experience. Therefore, the present of massive vehicle data in the vehicle networking ought to be efficient and humanized.

Many algorithms have been proposed to solve this problem, such as the aggregation algorithm on Baidu map makers [4], which includes the marker point aggregation, based on view rectangle and distance, and improved K-means algorithm based on Baidu map [5]. Currently, Google Maps provide a mark point aggregation service. Most map services do not provide a service based on the mark point aggregation API. At the same time, it is difficult to bear the mass of vehicle location data based on the client's mark point aggregation. Therefore, this paper focuses on the research of efficient marker aggregation algorithm.

In this paper, we propose a dynamic aggregation algorithm based on the administrative unit and a dynamic aggregation algorithm based on the K-means service mark point. The purpose of this paper is to find the best solution to the problem of map display of high load vehicle position data of the Vehicle Networking, so as to achieve efficient display, smooth interaction and friendly user experience. The high position of the vehicle dynamic load aggregation algorithm based on Vehicle Networking displays the location information of vehicles within user's limited map visual area. This approach shows the most comprehensive and accurate information without generating overlapping coverage. In addition, it optimizes the front-end display speed and avoids the browser collapse, and greatly improves the user experience.

The paper is divided into six sections. Section II gives a brief overview of previous work. Then in Section III we introduce the preliminaries of our work. Our method is presented in Section IV. In Section V we compare our proposed approach with the existing algorithms by experiments. Finally, conclusions are drawn in Section VI.

II. RELATED WORKS

There have been many efforts towards studying Vehicle Networking [1][2][3]. There are a lot of technical applications used in the vehicle networking [6]. Map service is one of the most popular applications [7]. Electronic map is widely used as a new technology [8]. Sun Jigui et al proposed the research on clustering algorithms [9]. For the map marking point aggregation, Zhou Ri et al proposed two clustering algorithms [10], the clustering algorithm based on grid and the clustering algorithm based on distance. Based on Baidu map makers, Dai Fengjiao et al proposed two polymerization methods, the marker point aggregation based on view rectangle and the marker point aggregation based on distance [4].

In order to improve the effect of polymerization, Yang Tingting et al proposed the method to improve the K-means algorithm based on Baidu map [5]. Liu Guo proposed an algorithm for massive spatial point aggregation display based on grid density [11].

III. PRELIMINARIES

The problem of map marking point aggregation has been studied extensively. Marking point aggregation, or tagging clustering, is one of the methods of map generalization, which mainly solves the problem of difficulty in expressing the midpoint of the map. And it can use a small number of points

978-1-5090-2508-4/17/$31.00 ©2017 IEEE

* Corresponding author

or icons to represent all points in the map, so that the map shows clearer.

At present, there are several kinds of mark point aggregation algorithms for online map: mark point aggregation algorithm base on grid, mark point aggregation algorithm based on distance, mark point aggregation algorithm based on the combination of grid and distance, and data mining clustering algorithm.

A. *Mark point aggregation algorithm based on grid*

The principle of Mark point aggregation algorithm based on grid is:

1. Divide the map into a square of the specified size.
2. Each zoom level corresponds to different sizes.
3. Then let the point in the corresponding grid be aggregated to the center of the square.
4. Finally, only one point is displayed in a square, and the number of primitive points contained in the aggregation point is displayed at the point.

This algorithm has the advantage of fast computation speed, since each original point is calculated only once without complex distance calculation. The time complexity of the algorithm is O(n). The disadvantage is that sometimes very similar points forced to open to different points of polymerization just because of the boundaries of the grid. In addition, the location of the polymerization point is the center of the grid, rather than the center of the grid, so that the aggregated points cannot accurately reflect the information of the original point. Polymerization effect is not good. Fig.1 shows a schematic diagram of grid based point aggregation algorithm before and after polymerization.

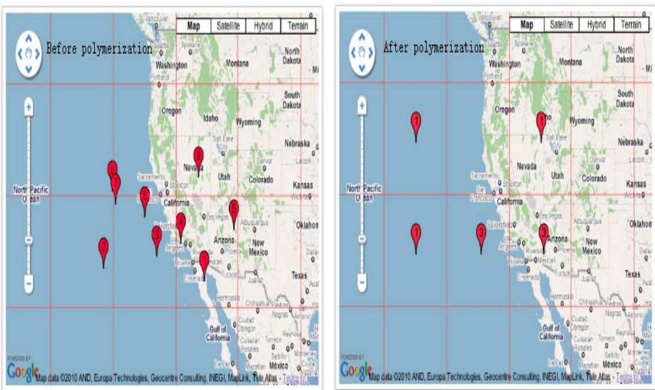


Fig.1 Schematic diagram of grid based aggregation algorithm before and after polymerization.

B. *Mark point aggregation algorithm based on distance*

The principle of Mark point aggregation algorithm based on distance is:

1. Aggregate based on the distance between points and points.
2. Traverse each point.
3. If the point of the traversal is within the range of the specified threshold of an existing aggregation point, so this point is aggregated to the point, otherwise create a new node, so cycle, until the end of the last point. The coordinates of the points after polymerization are the location of the first original point when the aggregation point is created.

The advantage of this algorithm is that the aggregation points accurately reflect the location information of the original points. The effect of polymerization is better. The disadvantage is that it needs to calculate the distance between points and aggregation points, the calculation is more complex. The time complexity of the algorithm is O (N^2). Fig.2 shows the aggregation process of a distance based mark point aggregation algorithm.

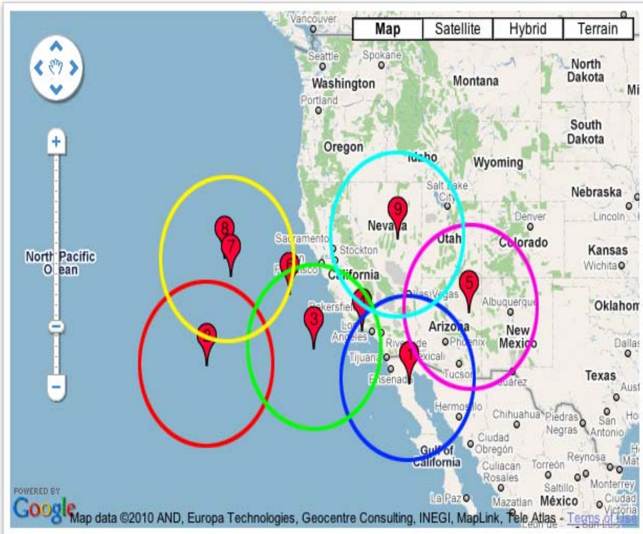


Fig.2 The aggregation process of distance based marking point aggregation algorithm.

C. *Mark point aggregation algorithm based on the combination of grid and distance*

The principle of Mark point aggregation algorithm based on the combination of grid and distance is:

1. There are no known aggregation points at the beginning.
2. Then iterate on each point, calculation of a point of outsourcing square.
3. If this point is not intersected with the outsourcing square of the existing aggregation point, then create new aggregation point.
4. If intersection, then the point polymerization to the aggregation point, such cycle until all points have been completed. Each zoom level iterates through all the original point elements.

The advantage of this algorithm is that the computation speed is fast, and the original points can be calculated only once, and the aggregation points can accurately reflect the position information of the original points. The disadvantage is that the algorithm is slower than the grid based aggregation algorithm. The time complexity of the algorithm is O(N^2). Fig.3 shows the aggregation process of the tag point aggregation algorithm based on grid and distance.

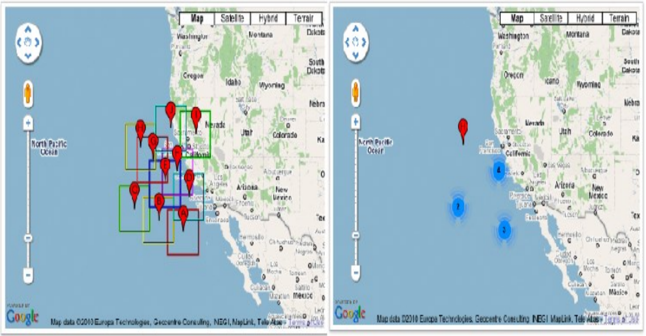


Fig.3 Aggregation process of mark point aggregation algorithm based on grid and distance.

D. Data mining clustering algorithm

The clustering algorithm of data mining can be divided into partition clustering algorithm, hierarchical clustering algorithm, density based clustering algorithm, grid based clustering algorithm, and other clustering algorithms.

The partition clustering algorithm needs to specify the number of clusters or cluster centers in advance. The error value of the objective function can be gradually reduced by iterative operation, and the final clustering result can be obtained when the objective function values converge. Partition clustering algorithm has a typical algorithm k-means algorithm [6] and k-medoids algorithm, etc.

Hierarchical clustering algorithm is also called tree clustering algorithm, it uses data association rules to separate or aggregate the data repeatedly and solves the clustering problem with a hierarchical sequence. The computational complexity of the algorithm is O (N^2). It is used to classification for small data sets [3]. Typical algorithms such as BIRCH(Balanced Iterative Reducing and Clustering Using Hierarchies), CURE(Clustering Using Representatives) and so on.

Clustering algorithm based on density is regarded as a region of high density in the data space. The main idea is that as long as the density of neighboring regions exceeds a certain threshold, the clustering is continued, data points in data sparse regions are considered as noise data. This method can be used to filter the "noise" outlier data and find any shape clustering. There are some typical algorithms such as DBSCAN (Density-based Spatial Clustering of Applications with Noise), OPTICS(Ordering Points to Identify the Clustering Structure) and so on.

The grid based clustering algorithm uses a multi-resolution grid data structure. It quantifies the space into a finite number of elements; the clustering operations are performed on the grid formed by these cells. The main advantage of this method is that the processing speed is fast, and the processing time is independent of the number of data objects, which is only related to the number of units of each dimension in the quantized space. There are some typical algorithms such as STING(Statistical Information Grid), Clique and so on.

IV. Improved Algorithms

According to the analysis the principle of all kinds of aggregation algorithms, and combined with the actual needs of mass data display, we propose two improved algorithms: The dynamic aggregation algorithm based on the administrative unit and the dynamic aggregation algorithm for K-means service mark point based on the administrative unit.

A. The dynamic aggregation algorithm based on the administrative unit

The principle of the algorithm is to map the layers of different zoom levels. We do dynamic aggregation of marker points in the field of view inside the national, provincial, municipal or county points. Taking county as an example, we take the county as the convergence point in the field of vision, traverse of each point in the field of view, and calculate the Euclidean distance of the points that are traversed to each aggregation point. Take the smallest aggregation point and add the points to be added to the aggregation point. So cycle, until all the points are finished. At the same time, this algorithm in the server running, that can be able to withstand massive data pressure. Dynamic aggregation refers to the movement or scaling of the map, according to the data in the field of view to re polymerization.

The main steps are as follows:

1. Enter the relevant parameters and define the relevant variables, Zoom represents the zoom level, Bound represents the field of view, Clusters represents the aggregation point, Vehicles represents the all vehicles.

2. According to the Zoom level, clean the data and get the corresponding administrative center aggregation points and the range of vehicle data. If the current Zoom level is greater than 12, return the vehicle data of the field of the view directly and display them.

3. Outer circulation traverses the range of each vision of the car, if the traversal is completed, return to the aggregation point data to show.

4. Inner loop iterates through each administrative center aggregation point, find the point where the vehicle is at a minimum distance, and add the vehicle to the aggregation point. The details of the algorithm are shown in Algorithm 1.

B. The dynamic aggregation algorithm for K-means service mark point based on the administrative unit

After designing the dynamic aggregation algorithm based on the administrative unit, it may not accurately reflect the information of the original point. So we produce the dynamic aggregation algorithm of K-means service mark point based on administrative unit.

The difference between the principle of the algorithm and the dynamic aggregation algorithm based on the administrative unit is that the former applies k-means algorithm in data mining clustering algorithm based on the latter. The performance of the former is that the latter aggregated K polymerization centers, the former calculates the center of mass of each aggregation center, and takes it as a new aggregation center, and then use the latter algorithm. Such iteration until the center of the polymer is no longer changed.

The front part of the algorithm is similar to the flow chart of the dynamic aggregation algorithm based on the administrative unit. Fig.5 shows only the iterative process of k-means. The main steps are as follows:

1. Take K administrative center in the view as the centroid.
2. Calculate the distance between the marker points in the field of view and the individual centroids, and each marker point is clustered with the nearest centroid.
3. To judge whether a marker points belonging to a cluster change. If not, the algorithm ends. Otherwise, take the average of all points in each cluster as the new centroid. Repeat the steps 2 and 3.

The detail is shown in Algorithm 2.

Algorithm 1: The dynamic aggregation algorithm based on the administrative unit

```
begin
  input Zoom; Bound; Clusters; Vehicles;
  if Zoom < 12:
     if Zoom < 9:
        if Zoom < 7:
           if Zoom < 4:
              Take the vehicles in both national level administrative center aggregation point and the field of vision;
           else:
              Take the vehicles in both provincial level administrative center aggregation point and the field of vision;
        else:
           Take the vehicles in both municipal level; administrative center aggregation point and the field of vision;
     else:
        Take the vehicles in both county level; administrative center aggregation point and the field of vision
     The Cluster  and Vehicles after the Zoom level, define the variable i=0, j=0, k=0, Max=Double.Max, Min;
      for i ← 0 to all vehicles:
         Min = Max;
         for j ← 0 all aggregation points:
           Calculate the distance between the first i and the first j aggregation(Dis);
              if Min > Dis:
                    Min = Dis, record the current aggregation point ,k=j;
              else:
                    repeat;
         end
         add the first i vehicle to its nearest aggregation point;
         repeat;
      end
      return the administrative center aggregation point;
  else:
     return the vehicles in the visual;
end
```

Algorithm 2: The dynamic aggregation algorithm for K-means service mark point based on the administrative unit

```
begin
  input Zoom; Bound; Clusters; Vehicles;
  ...
  take the k administrative center in the view as the center of mass
  calculate the distance between the marked points in the field of view and the center of the mass;
     if all the clusters of marker points did not change:
        return;
        end
     else:
         take the average of all the points in each cluster as a new centroid;
end
```

V. Experimental Verification

In order to In get better results, we compare our two algorithms and compared with other algorithms. These algorithms run in the computer which is 3.60GHz frequency, quad core processor, graphics card, 1G memory, 8G RAM, 1T capacity and speed of 7200 RPM hard drive, the operating system is Windows10. The experiment was carried out by using Google Chrome browser.

A. *Experimental object*

The experimental object is the dynamic aggregation algorithm based on the administrative unit and the dynamic aggregation algorithm for K-means service mark point based on the administrative unit and the aggregation algorithm provided by Baidu map. If there is no special instruction, we use A to represent the Baidu map marker point aggregation algorithm, B represents the dynamic aggregation algorithm based on the administrative unit, and C represents the dynamic aggregation algorithm for K-means service mark point based on the administrative unit.

B. *Experimental data sets*

1. Vehicle location data. Vehicle location data is real which is from a company vehicle networking project. In this paper, the location data of 10000 vehicles were extracted randomly.
2. The latitude and longitude data of administrative unit. The algorithms proposed in this paper need to use the latitude and longitude data of province, cities and counties. In this paper, we use the data in 2015.

TABLE I The corresponding administrative units for different scaling levels

Zoom level	Aggregation point
3,4	Country
5,6,7	Province
8,9	City
10,11,12	County

C. *Experimental parameter setting*

In this paper, the scale of the algorithm Zoom is [3]. Different scaling levels use different administrative units as aggregation points, as shown in TABLE I. The value of K in the algorithm C takes the number of administrative units in the range of the current map.

D. *Experimental results*

We mainly from the response time, memory footprint, polymerization effect to compare the results of the experiment.

1. The response time

In order to reduce the experimental error, the response time was 1000，2000... 10000. The calculation of each response time is from the start of the map zoom level 3 to the end of 12. And then calculate the average response time of each scale aggregation.

TABLE II shows the average response time of the three algorithms for different vehicles.

TABLE II The average response time of the three algorithms for different vehicles

Agorithm	Average response time (MS) for different number of vehicles			
	1000	2000	...	10000
A	297. 3	571. 9	. . .	2657. 4
B	32. 8	39. 4	. . .	38. 1
C	34. 5	42. 4	. . .	65. 1

In order to compare three kinds of algorithms using the aggregation response time more intuitively, we also adopted the representation method of histogram.

Fig.4 shows the influence time histogram of the three algorithms.

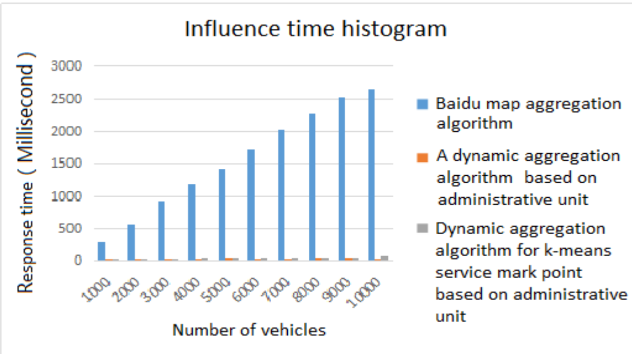


Fig.4 The influence time histogram of the three algorithms

2. The memory footprint

This paper uses the Google Chrome browser to carry on the experiment. We use the Chrome own task manager to calculate the memory footprint. In order to get a more accurate calculation results, for each algorithm in the map from the zoom level 3 to level 12, then gradually reduced to the level of 3. We can determine the final memory usage to the algorithm until the Chrome task manager displays the memory footprint stably.

The experiment was carried out respectively in the number of vehicles at 1000, 2000 ... 10000.

TABLE III shows the memory footprint after polymerization for different number of vehicles.

TABLE III The memory footprint after polymerization for different number of vehicles

Algorithm	The memory footprint in the browser for different number of vehicles (megabytes)			
	1000	2000	...	10000
A	96	103. 1	. . .	137. 2
B	67	69	. . .	88. 4
C	66. 2	69. 6	. . .	87

In order to comparison the memory footprint after polymerization of three kinds of algorithms more intuitive, we also adopted the representation method of histogram. Fig.5 shows the memory footprint after polymerization histogram of the three algorithms.

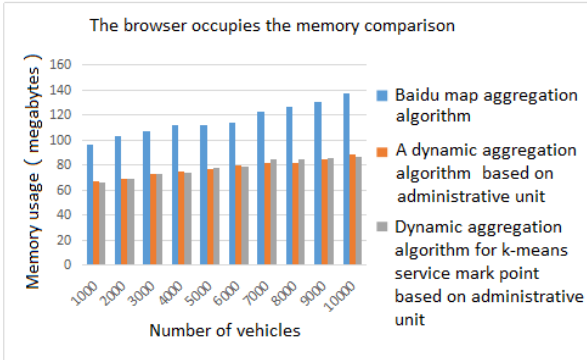
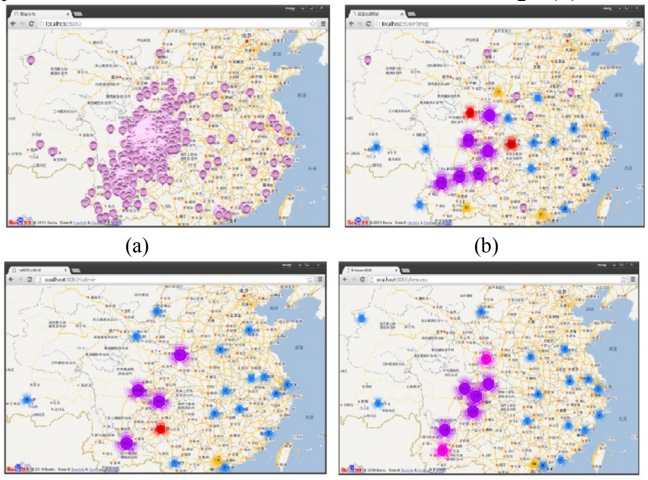


Fig.5 The memory footprint after polymerization histogram of the three algorithms

3. The polymerization effect

The comparison of the polymerization was carried out under the number of vehicles at 2000. First, we recorded the original vehicle distribution such as Fig.6(a), then we recorded the aggregation results of Baidu map marker point aggregation algorithm such as Fig.6(b), the dynamic aggregation algorithm based on the administrative unit such as Fig.6(c) and the dynamic aggregation algorithm for K-means service mark point based on the administrative unit such as Fig.6(d).



(a) (b)

(c) (d)

Fig.6 The polymerization effect

VI. CONCLUSIONS

In this paper, we make the experiments by real vehicle data which provided by the company. We compared the effects of the three algorithms in response time, memory usage, and polymerization effect. The experimental results show that the response time and the memory usage of Baidu map marker point aggregation algorithm is more than others', and its dynamic polymerization effect is poor. The dynamic aggregation algorithm based on the administrative unit has the less response time and memory usage, and the dynamic polymerization effect is the best. The dynamic aggregation algorithm for K-means service mark point based on the administrative unit also has the less response time and memory usage, and its dynamic polymerization is better than Baidu map marker point aggregation algorithm but worse than the dynamic aggregation algorithm based on the administrative unit. The experiments show that the two algorithms proposed in this paper have better performance than the state-of-the-art methods.

ACKNOWLEDGMENT

This work was supported by Science and Technology Guiding Project of Fujian Province, China (2015H0037, 2016H0035); NSFC (No.61402387); NSFC (No.61402390); Science and Technology Key Project of Fujian Province, China (2014H0044); Education and Research Project of Middle and Young Teacher of Fujian Province, China (No.JA15018, No.JAT160003); Science and Technology Project of Xiamen, China (3502Z20153026).

REFERENCES

[1] Hu Xinyu, Zhang Jie. The development and challenge of the vehicle networking[J]. Internet of things Technologies, 2017(02): 56-59.

[2] Su Jing, Wang Dong, Zhang Feifei. Summary of Application Technology of vehicle networking Technology [J]. Internet of Things Technology, 2014 (6): 69-72.

[3] Zheng Zhi, Wei Aiguo, Gao Wenwei. Vehicle networking technology and development [J]. Journal of Military Communications College, 2014 (3): 70-73.

[4] Dai Fengjiao, Xiao Linhua, Yanglu, Liu Zhiyuan. The research on the algorithm of mark point aggregation based on baidu map[J]. China Science and Technology Information, 2013(23): 82-85.

[5] Yang Tingting, Wang Xuemei. The research on improved K-means algorithm based on Baidu map[J]. Computer Engineering & Software, 2016(1): 76-80.

[6] Li Zhihan, Liu Tingrang, Guo Zhongyi. Summary of technology application of the vehicle networking[J]. Science and Technology, 2015(12): 100-100.

[7] Zhao Juan, Liu Wenjie. A new chapter of open the map service[N]. The newspaper of chinese Surveying, 2016-01-29(003).

[8] Wei Hongyan, Chang Jun. Analysis on the present situation and application prospect of electronic map[J]. Intelligence, 2015(10): 320-320.

[9] Sun Jigui, Liu Jie, Zhao Lianyu. Research on clustering algorithm [J]. Software Institute of Chinese Academy of Sciences, 2008(19): 48-61.

[10] Zhou Ri, Miao Fang. The polymerization of electronic map marker[J]. Electronic Technology & software Engineering, 2014(6): 158-159.

[11] Liu Guo. An algorithm for massive spatial point aggregation display based on grid density[J]. Geomatics & Spatial Infor mation Technology, 2015(4): 174-176.